# Innate and Adaptive: A Combined Arms Approach to Artificial Immunity

Hugo Robinson, 690065365

December 2022

### Abstract

In this paper a critical review of The Human Immune System, Artificial Immune Systems and Clustering Algorithms is presented in order to develop a Intrusion Detection System that combines the innate and adaptive immune system in a novel way. Due to the increasing complexity and difficulty to detect zero-day attacks, classic Intrusion Detection Systems require support from novel approaches. This, as with many other areas of computer science, can come from nature. Artificial Immune Systems are a relatively recent topic for computer science, various discussions in immunology around how the Human Immune System works have allowed multiple methods which are based on different theories to be proposed. The two major ideas discussed include the ability for the body to distinguish self from non-self, which takes the form of individual immune cells scanning for pathogens within the body, or that tissue cells emit signals to the immune system when they are damaged or killed which supports a more distributed approach to the immune system. Clustering algorithms are better researched, however specific traits are required for this project, including the unsupervised changing of the number of clusters it produces over its lifespan. This is usually seen in algorithms that include neural networks as a core part of its functionality, or modified classic algorithms, presenting both advantages and disadvantages to using either. Further to this, a project specification is presented, detailing the projects objectives and providing a timeline. Testing and validation criteria have been laid out in line with validation techniques of machine learning algorithms as the system as a whole will attempt to learn and cluster new attacks.

*I certify that all material in this dissertation which is not my own work has been identified.*

# Contents

# 1 Introduction

With 39% of businesses in the UK reporting cyber attacks in 2022 [1] and the cyber security sector reported to grow from USD 155.83 billion in 2022 to USD 376.32 billion by 2029 [2] cyber security is a large part of modern day infrastructure [3], [4]. Therefore, secure systems should be in place to protect our vulnerable infrastructure, however classic cyber security methods, namely intrusion detection, are not perfect. Issues can arrive with high false positive rates and difficulty adapting to new threats as their database must be updated by hand [5]. However as with many problems in computer science, nature may help. Providing a system that is adaptable, resilient and distributed, the Human Immune System (HIS). The HIS already demonstrates many of the required characteristics of an effective Intrusion Detection System (IDS) and has recently been studied and adapted to help with the issues mentioned earlier. However, due to different theories within the field of immunology itself [6] there is no one solution to the problem creating the novel, yet expansive field of Artificial Immune Systems (AIS).

The first part of this paper will discuss the current literature around the Human Immune System, describing its functionality and current theory's. This is in order to provide a basis of understanding for the second part, which will further look at current Artificial Immune Systems research and how it draws from different HIS. The third part of the literature review will discuss clustering algorithms, this forms the second half of the project allowing the system to both detect and classify attacks. Finally is a project specification which will outline the aims, objectives and timelines for the completion of the project.

# 2 Motivation and Research Context

## 2.1 Description of Research Problem

The ability to support current anti-virus detection systems with a dynamic intrusion detection system would allow further security enhancements across networks without a large overhead, in turn increasing protection and not slowing down communication. However to implement such a system, certain challenges must be overcome.

Firstly, the network communications must be transformed into data that can be parsed quickly by an AIS such as a binary representation. However this cannot cause too much overhead or the networks latency will be increased.

Secondly, once the network communication has been transformed it must be quickly parsed by the AIS to identify any anomalous communications. There are a number of key areas of the AIS that must be implemented, the threat detection must be fast, the AIS must be able to distinguishes between normal and abnormal and the AIS must be able to be updated if necessary.

Finally, when an abnormal communication is detected an ML classifier is used to identify what the communication could be, based on previous attacks, much like traditional anti-virus software. This part must be able to identify attacks which are similar to each other, whilst also creating new classes if there is an unknown attack. This information can either trigger an automatic response or provide a report to a human operator.

To narrow the scope of this problem certain assumptions will have to be made and further areas of the project will not be developed, such as automatic responses and simulated networks.

## 2.2 Zero-Day Attacks

A zero-day attack is an attack on a system that has not been seen before and therefore is potentially not protected against by classic intrusion detection software [7]. This is because classic intrusion detection software matches simple patterns with a know database of attack patterns [8], therefore if the pattern is not in the database the attack may go undetected until serious damage is dealt. A system must be developed that is able to detect anomalous interactions that could be zero-day attacks.

## 2.3 Human Immune Systems

The human immune system has struggled with the problems now presented to computer security for thousands of years, evolving alongside our other internal systems [9]. How can an enclosed system define what is itself and what is not with the ability to distinguish this?

Intuitively the immune system seems best suited to detect self from non-self and was the first proposed immunological model [10], where the body knows what should be there and what shouldn't. However in the paper [11], Matzinger discusses a different approach to how the immune system initiates a response. Where the immune system actually distinguishes dangerous and safe instead. The paper goes on to propose that instead of immune cells coming across a pathogen and activating an immune response it is instead tissue cells that will emit danger signals when damaged or have died for reasons other than from apoptosis. This initial danger response is part of the innate immune system and is vital for an adaptive response [12]. Both of these systems are discussed further.

There are a number of different cells found within the body that support the immune system in various ways and can be split into two categories: Innate and Adaptive [13]. The innate immune system is a non-specific immune response where the body can rapidly initiate a response to general pathogens that share common structures. Such responses can take the form of inflammation, an increase in temperature and specialised white blood cells flooding the site. This is achieved through pattern recognition where similar pathogens can be detected and a rapid response can be deployed. These are known as pathogen associated molecular patterns (PAMPs)[14] and provide a possible way for computers to mimic a similar function for cyber security, as the innate system is fast to respond using pre-encoded receptors on the germline [15]. Part of the innate immune system is the complement immune system, this works by distributing a very large number of different Antigen-presenting cells (APCs) throughout the body, when these APCs interact with a pathogen they activate and cause an influx of other proteins, inflammatory events and can mark pathogens for destruction by white blood cells [16]. An APCs will achieve this through the Pattern Recognition Receptors (PRRs) found on itself. These receptors receive either the previously discusses PAMPs or if the issue is a host cell, Damage-Associated Molecular Patterns (DAMPs) [17]. These APCs not only act as a pathogen identifier, but also help to activate the adaptive immune system by releasing signals to native T-Cells further activating them and allowing the adaptive immune system to start[18]. This early identification, support and adaptive activation is vital for defending against more dangerous infections.

The second system is the adaptive immune system, a targeted response to a specific pathogen. As the adaptive immune systems is particularly dangerous to anything that is deemed mom-self, it must be able to differentiate between the two [19], because if this system fails and the cannot differentiate between self and non-self then the immune system will attack itself or the host, this is autoimmune disease [20]. As previously mentioned the first active component of the adaptive immune system are the T-cells, once activated by their specific APC they may bind and kill infected cells, regulate immune reaction or stimulate B-Cells [21]. These B-Cells then produce specific antibodies to combat the specific pathogen encountered. A subset of produced B and T cells become memory cells which specialise on an encounter with a pathogen but are not activated during the initial immune response. These memory cells, once specialised, remain within the system longer than regular B and T cells meaning if reinfection occurs a faster adaptive immune response can be mounted as more cells are already specialised [22].

What has been previously discussed is a classical view of the immune system, however recently a new school of thought has emerged which provides a new and interesting perspective on the immune system which may provide a new way to develop Intrusion detection Systems. This theory is Danger Theory [11], as the immune system is only able to detect pathogens occasionally, Danger Theory proposes that it is not that the immune systems detects between self and non-self, but reacts to a number of danger signals released by affected cells within the body. Therefore it is not an attempt to distinguish a foreign invader, but a dangerous entity that damages the body. It is suggested that as a pathogen causes cell stress or cell death, these cells release molecules that are recognised by the APC's discussed above as part of the innate immune system, which in turn trigger the adaptive immune system. These danger signals may not be released only by cells within the body, but may also be released by the invading pathogen [23] which also trigger the APC's in a more traditional way.

The human immune system is complex and not fully understood however certain properties can be used to develop computer systems with similar effects.

## 2.4 Artificial Immune Systems

As computer networks have grown and individual devices have become intertwined it has become more paramount to detect unauthorised access and abnormal behaviour within these systems. One of the major components of this problem are Intrusion Detection Systems (IDS) first idised in 1972 [24] and developed upon in 1980 [25]. Throughout the 1980's and 1990's further models for IDS were developed, including the Intrusion Detection Expert System (IDES) [26] and the Network Security Monitor (NSM) [27]. The details of the human immune system are discussed above and, in a long tradition of using natural processes in computer science, has been used to develop these IDS [28], with the first mention of AIS in 1986 [29]. With AIS being used in many different scenarios and having many different iterations [30]–[32], the literature must be assessed to decipher the best one for this project.

Research on AIS has split into four main algorithms:

### 2.4.1 Negative Selection Algorithms

NSAs take the idea of certain detectors that encompass an area of non-self, if these detectors match with something it can be classed as non-self [33]. The process begins with random detectors being generated within the search space, some of which will overlap the self-space. These detectors that overlap will be culled as to not flag a false positive by detecting the self-space, this is similar to a process that happens when white blood cells mature in the thymus [34]. The detectors usually work by comparing incoming data to themselves in the form of binary strings. Although this approach works, there are a number of drawbacks; some problems require a large number of detectors to be generated which could prove incompatible in some cases. The use of binary strings reduces the domain knowledge, therefore making it much harder to decipher a reason for detection. [35]. In its early form the detection of non-self was not fuzzy and therefore not representing the true state that normalcy is usually fuzzy, although this is addressed by Mr Silva et al [36]. Further to this Gonzalez et al [37] lays out a NSA using real values (RNS) to solve some of these problems and increase the processing time by using real values instead of a binary representation.

### 2.4.2 Colonal Selection Algortihms

CSLs are a family of genetic algorithms that are based on the idea that within the body, only cells that can identify a pathogen will continue to proliferate [38]. De Castro and Von Zuben [39] proposed the system CLONALG that takes ideas from genetic algorithms by creating an initial population of antigens, used to detect anomalies. Then evaluating these antigens and only allowing the best to move onto the next generation and mutate. Similar to traditional genetic algorithms there are a wide range of differences that can be applied to the base algorithm such as changing the sampling strategies [40]. When compared to other heuristic algorithms such as Ant Colony Optimization and normal genetic algorithms, CSA seems to perform better [41].

### 2.4.3 Immune Network Algorithms

Immune network algorithms take inspiration from neural networks, where detectors within the system are connected via edges that dictate similarity to one another. This is because the stimulation strength of the B cells within the body that Jhon Timmes et al [42] attempt to simulate, relates to both how the B cell binds with the antigen and its enmity to its neighbouring B cells [29]. As the B cell detector is influenced by both the pathogen and its neighbours a strong response can be suppressed by having fewer or distant neighbours. This will stop the B cell from reaching a stimulation threshold and cloning itself. Although these cloned cells will undergo some mutation, producing a cluster of cells that are similar, allowing the system to quickly detect similar infections. This technique appears to be an effective clustering tool, however as the required data set grows, the depth of the system will grow as well, even if some weaker cells are culled [43].

### 2.4.4 Danger Theory Algorithm

As Danger theory algorithms take inspiration from danger theory they take a different approach to previously discussed algorithms. As discussed in the paper [44] a danger theory intrusion detection system is based on how the system interacts with varying types of alerts. It goes on to discuss apoptotic and necrotic alerts, these represent normal cell death, where a cell is either instructed to die by the body, or does it of its own accord [45]. In a computing sense, this represents a low level alert or, routine maintenance so, unless there are a large number of these alerts it is not necessary to trigger an immune response. On the other hand necrotic cell death occurs when a cell is destroyed outside of the regulated pathways , usually caused by a pathogen or intruder [46]. This translates to detecting unexpected outcomes from the system running compared to normal. The paper provides an excellent example based on a DDOS attack, where an attacker just scanning the system would trigger an apoptotic alert, however actual system damaged caused by the attack would trigger a necrotic alert.

Some interesting further research to this may involve developing local danger alerts where a network of intrusion detection alerts are triggered throughout a system, allowing the IDS to both mitigate the response to the affected areas and provide a detailed report on the scale of the attack.

## 2.5 Machine learning for Classification

The work previously discussed has dealt with the aspect of intrusion detection. However this only simulates the innate immune system, to develop a well rounded AIS the adaptive immune system must also play a role. Therefore attacks must be classified and remembered to allow for a more effective response if the attack were to happen again. Clustering algorithms are an excellent way to do this, as similar attacks will display similar properties to each other, allowing for clusters of attacks to form and be identified, so when a new, similar attack is launched against the system a faster response can be initiated.

There are a large amount of clustering algorithms [47], however here only some will be considered.

### 2.5.1 K-Means

K-Means is one of the most popular clustering algorithms. The algorithm is initialised with a known K value. This K value determines how many clusters the algorithm will generate for the data set. Determining these clusters is an iterative process where the algorithm will randomly choose a K number of points which become the center of the clusters. After this each other data point is compared to the center of the clusters and is assigned to the closest one. However this is not the optimal solution, therefore the center of the clusters are moved to be an average of all of the points assigned to it. This process repeats itself until the the variance in the clusters no longer changes [48].

There are some disadvantages to K-Means, it has trouble clustering data of varying sizes, which may be the case when it comes to network attacks. Equally it must know the hyper-parameter k beforehand which is unknown to the user.

### 2.5.2 Density Based

Density based clustering algorithms such as DBSCAN work by detecting areas where there is a concentration of data points and areas of sparsity [50]. The DBSCAN algorithm has two user set parameters and identifies three types of points. The first parameter is the radius of the circle around

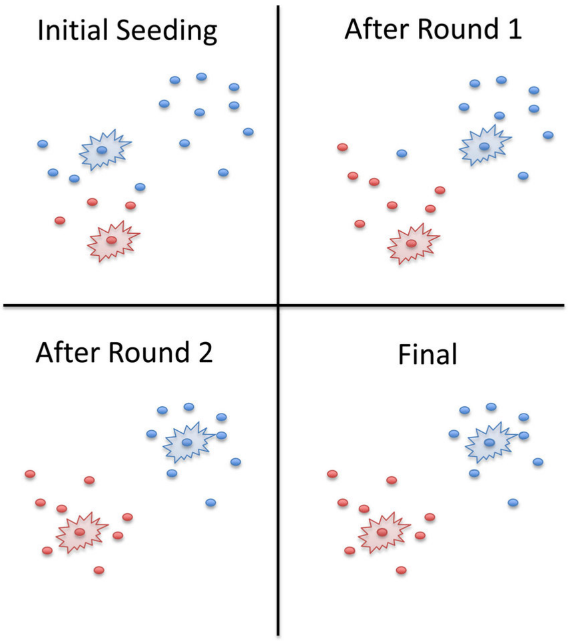

Figure 1: This image shows an example of K-Means clustering where K=2. The clusters are initially set to two data points, but are then reassigned to the average of all of the points in their clusters [49]

each point that defines its neighbourhood, the second is the threshold to become a core point. A core point is a point that has a set number of other points around it, a non-core point is a point that does not reach this threshold of neighbours and an outlier is a point with no neighbours that are core points. The algorithm checks every point and labels them as core, non-core or an outlier. One random point is then chosen to become the first cluster and all core and non core points in its neighbourhood are added. This cascades through the cluster until it is complete. Then another core point is chosen to start a new cluster. This continues until all non outlier points are added to a cluster [51].

Density based clustering has the distinct advantage of being able to identify nested clusters, unlike K-Means. Also DBSCAN does not require the number of clusters to be specified, allowing for unknown data to be clustered. Finally it is robust to outliers and noise within the data as they will not affect the clusters. However there are some disadvantages to DBSCAN, if the data set is too sparse then some similar points may not be clustered together. It is also heavily reliant on the starting parameters to be effective [52].

### 2.5.3 Hierarchical

Hierarchical clustering provides a more unique way to cluster data sets, especially data sets that are hierarchical in nature. This is because it produces a tree of clusters where cutting the tree in different places will generate a different precision of clusters. Hierarchical clustering works by comparing points to each other and combing points into groups, creating a dendrogram. The first point is compared to all the others to find the most similar and this pair is combined into a group. This continues for all of the data points which are either paired together or added to a group. This process continues until there is one group containing all of the data points, as shown in figure 2.

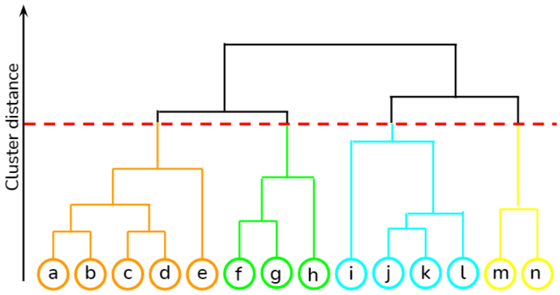

Figure 2: This image shows an example of Hierarchical Clustering, showing the different groups form over time [53]

As can be seen in figure 2 there is a red line denoting the cut off point and the clusters below this line can be seen in different colours. If the line were to be moved up or down, different clusters could be shown.

Hierarchical clustering provides some benefits such as a reduced sensitivity to noise, no requirement for a number of clusters parameter and most importantly, the ability to increase or reduce the accuracy of the clusters. However there are some major disadvantages to hierarchical clustering, such as it handles convex shaped clusters poorly and it is not scalable for large data sets [52].

### 2.5.4 Self Organising Maps

Self Organising Maps are a form of Artificial Neural Network that uses neurons and weights to cluster data. For self organising maps, a number of nodes in a one or two dimensional space are connected via weights to data points in a n-dimensional space, the neurons are also connected to each other. These neurons are distributed randomly around the search space. For every data point a ”winner neuron” is calculated, which is the neuron that is closest to the data point. Once this is calculated the neuron’s weights are updated and the neuron will move closer to the data point. As all of the neurons are connected all of the other neurons are also moved slightly. Once all of the data points have been calculated each neuron becomes a cluster and the all the data points are assigned to their closest cluster within a set hyper-parameter distance.

Self organising maps have the distinct advantage of being both a clustering algorithm and reducing the dimensions of the data from n-dimensions to 2-dimensions. They have also been shown to classify data up to 10 times faster than other methods such as K-Means. This is due to the fact that a new map does not need to be created every time a new piece of data is added. However there are some disadvantages to Self Organising Maps, namely that they can take a longer time to train, therefore if the map was lost during an attack it would be difficult to retrain it quickly [54].

# 3 Conclusion of Literature Review

Overall the literature provides both some more insight into the topic and some questions to be answered. With the immunological community unsure as to which theory is more accurate it provides a larger scope for different implementations of AIS for computer scientists. With Negative selection, Colonal Selection and Immune Networks all taking a similar approach to each other based around the differentiation of self and non-self whereas Danger Theory takes the approach of cells sending out signals when they are damaged, alerting the system.

Further to this, different clustering algorithms were looked at to support the creation of an adaptive immune system. However multiple of these algorithms require a base understanding of how many clusters there are before it runs and are set as hyper-parameters. This is unsuitable for the proposed project as there are an unknown number of possible zero-day attacks and therefore the number of clusters cannot be known beforehand, therefore algorithms such as Self Organising Maps and other neuron based algorithms appear to be better suited to the problem.

# 4 Project Specification

## 4.1 Project Aim

This project will research and build a system that is both an innate and adaptive immune system, this is achieved through a detection system (innate) that decides what is self and non-self / danger quickly from network messages. If a message is flagged, it is sent to a clustering algorithm that helps to identify what it is (adaptive). This, in future, would allow a quick targeted response to similar threats, much like human immune systems. To link it to human immune systems the project will be an AIS that can identify ”pathogens” (Attacks) in order to create ”antigens” (Countermeasures). This is in order to support a quick and effective response to zero-day attacks.

## 4.2 Project Objectives

This project will be split into five major stages:

The first stage will focus on identifying a suitable data set. A suitable data set will take the form of pre-made network communications that already contain some labeled attacks. This is required as the first stage as the rest of the project will be tested and built using the data set. If this data set does not exist or cannot be found, the contingency plan is to create a data set of binary strings which all follow a distribution of a few base strings. The attacks will be simulated through a distribution of different strings/outliers from the original distribution. This stage is to be complete no later than 9th January.

The second stage is to wrangle the data as it may not be in an appropriate form to work with. This may involve converting all of the messages into binary representations and splitting the attacks out from the normal communications in order to further train the models. The contingency plan for this stage is the same as stage one. This stage is to be complete no later than 16th January.

Stage three is the development of the AIS itself. As the data set is already in place it can be used to test against. Completion of this stage will see a single node of a distributed, danger theory based AIS implemented that is able to flag danger signals to identify unknown messages. Due to other commitments this is to be complete no later than 27th February. As this is a major part of the project if the deadline is not hit it must push into the timeline of stage four, therefore stage four will push into stage five and stage five will reduce in scope.

Stage four will see the development of the adaptive immune system through the implementation of a clustering algorithm. Completion of this stage will see the implementation of an unsupervised clustering algorithm that can generate its own clusters as it sees fit from the data provided by the AIS. The algorithm must show that it can cluster similar messages together, providing a reasonable capability to distinguish between different attacks. This is to be completed no later than 27th March. As this is a major part of the project if the deadline is not hit it must push into the timeline of stage five, therefore stage five will reduce in scope.

Stage five is the final stage, where the model will undergo testing and validation. Performance measures will be run against the system to test for false positives, false negatives and true positives. This is in order to test the effectiveness of the system and provided data for further improvements. This is to be complete no later than 17th April.

If these goals are completed within the time frame the stretch goals are to simulate two nodes that can flag danger to each other, this will provide a more comprehensive response from the system as two danger signals will be flagged. Also an output showing the clustering in a reduced dimensional space that a human operator can read to help determine if something is a threat or to be used in a report. These objectives do not have a time scale as they are stretch goals.

## 4.3 Project Management

### 4.3.1 Risks

The project does pose a number of risks:

Due to the novel nature of the topic, relatively few papers have been published. This reduces the pool of available knowledge of implementation and troubleshooting to rely on throughout the project. Therefore before each stage a defined plan must be set, based on the results of the previous stage.

As discussed above, one major risk is that no data set is suitable for the project, therefore the data set must be self generated. This will be accomplished by generating binary messages using a distribution and using far outliers as attacks that are similar to messages and other specially curated messages that are very different to the normal ones to simulate more obvious attacks. These attack messages will also follow a distribution to allow for clustering to take place.

### 4.3.2 Legal and Ethical

All legal matters will be considered when using third party data sets. No paid for data sets will be used. Any data that is identifiable will be obfuscated before storage and use.

The nature of this project means there are no ethical risks as it will not be processing any identifiable data or deploying malicious software and attacks. Any attacks seen within data sets will not be run as it is kept internal and not sent over a network.

### 4.3.3 Gantt Chart

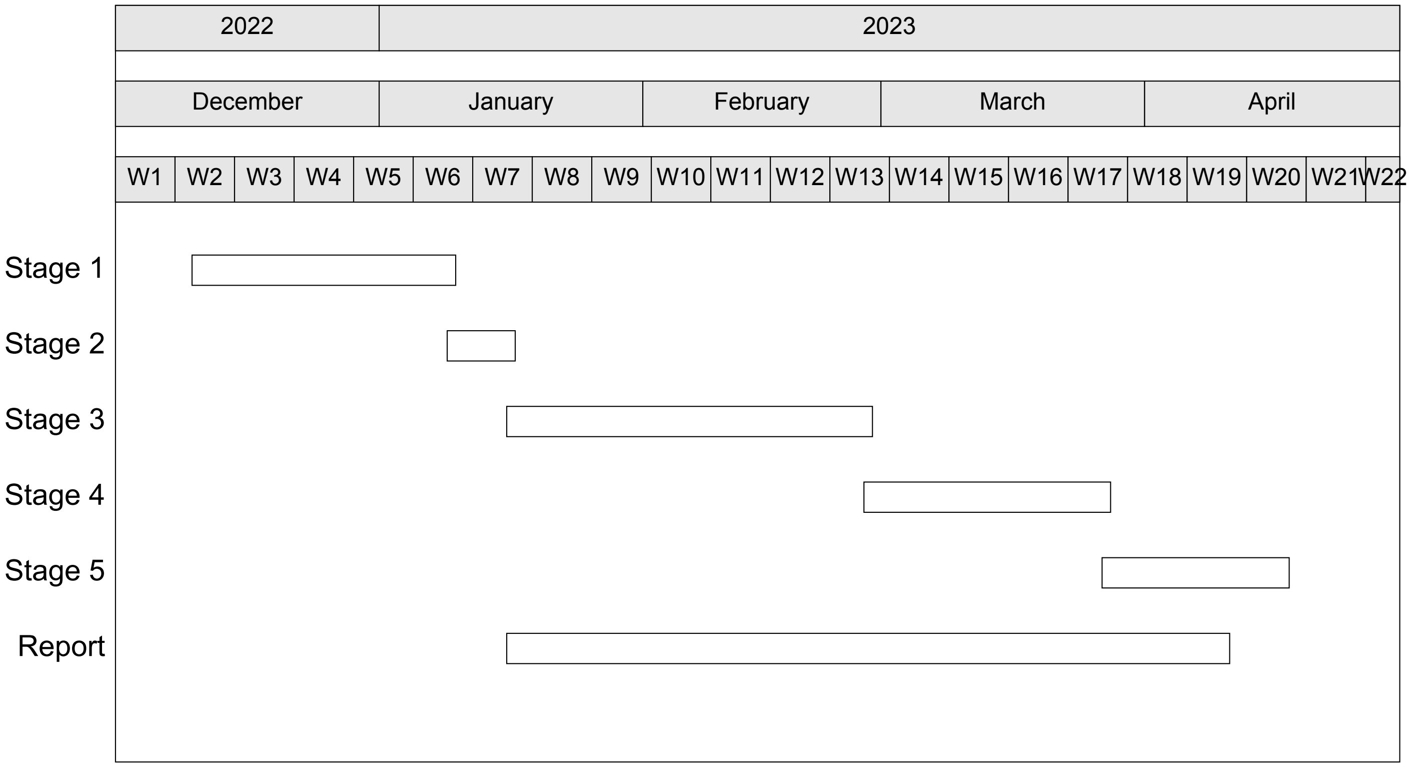

Figure 3: This Gantt chart shows the project timeline and duration of each stage

## 5 Conclusion

Overall the project will consist of an innate immune segment, characterised by an AIS conducting ID on network communications. If an intrusion is detected it will pass the communication to the adaptive segment, which is a clustering algorithm, to attempt to identify what the possible threat is based on previous data. With the use of the previously discussed literature, this section presents a well scoped project plan, with defined and time bound objectives. Contingency plans for possible complications, and the identification of risk and legal and ethical issues.

# 6 Bibliography

## References

[1] *Cyber security breaches survey 2022*. [Online]. Available: https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2022/cyber-security-breaches-survey-2022 (visited on Nov. 29, 2022).

[2] *Cyber security market overview by size, growth amp; trends, 2029*. [Online]. Available: https://www.fortunebusinessinsights.com/industry-reports/cyber-security-market-101165 (visited on Nov. 29, 2022).

[3] J. Peterson, M. Haney, and R. Borrelli, "An overview of methodologies for cybersecurity vulnerability assessments conducted in nuclear power plants," *Nuclear Engineering and Design*, vol. 346, pp. 75–84, 2019.

[4] A. O'dowd, *Major global cyber-attack hits nhs and delays treatment*, 2017.

[5] M. Aljanabi, M. A. Ismail, and A. Ali, "Intrusion detection systems, issues, challenges, and needs," *International Journal of Computational Intelligence Systems*, vol. 14, Jan. 2021.

[6] J. Köhl, "Self, non-self, and danger: A complementary view," *Current topics in complement*, pp. 71–94, 2006.

[7] L. Bilge and T. Dumitraş, "Before we knew it: An empirical study of zero-day attacks in the real world," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ser. CCS '12, Raleigh, North Carolina, USA: Association for Computing Machinery, 2012, pp. 833–844.

[8] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[9] G. Hemmrich, D. J. Miller, and T. C. Bosch, "The evolution of immunity: A low-life perspective," *Trends in Immunology*, vol. 28, no. 10, pp. 449–454, Oct. 2007.

[10] F. M. Burnet *et al.*, "The production of antibodies. a review and a theoretical discussion.," *The Production of Antibodies. A Review and a Theoretical Discussion.*, 1941.

[11] P. Matzinger *et al.*, "Tolerance, danger, and the extended family," *Annual review of immunology*, vol. 12, no. 1, pp. 991–1045, 1994.

[12] P. Matzinger, "The danger model: A renewed sense of self," *science*, vol. 296, no. 5566, pp. 301–305, 2002.

[13] M. F. Flajnik and M. Kasahara, "Origin and evolution of the adaptive immune system: Genetic events and selective pressures," *Nature reviews. Genetics*, vol. 11, no. 1, pp. 47–59, Jan. 2010.

[14] J. S. Marshall, R. Warrington, W. Watson, and H. L. Kim, "An introduction to immunology and immunopathology," *Allergy, Asthma & Clinical Immunology*, vol. 14, no. 2, p. 49, Sep. 2018.

[15] R. Medzhitov and C. A. Janeway Jr, "How does the immune system distinguish self from nonself?" en, *Seminars in Immunology*, vol. 12, no. 3, pp. 185–188, Jun. 2000.

[16] J. Charles A Janeway, P. Travers, M. Walport, and M. J. Shlomchik, "The complement system and innate immunity," *Immunobiology: The Immune System in Health and Disease. 5th edition*, 2001.

[17] G. P. Amarante-Mendes, S. Adjemian, L. M. Branco, L. C. Zanetti, R. Weinlich, and K. R. Bortoluci, "Pattern Recognition Receptors and the Host Cell Death Molecular Machinery," *Frontiers in Immunology*, vol. 9, 2018.

[18] S. J. Gaudino and P. Kumar, "Cross-Talk Between Antigen Presenting Cells and T Cells Impacts Intestinal Homeostasis, Bacterial Infections, and Tumorigenesis," *Frontiers in Immunology*, vol. 10, p. 360, Mar. 2019.

[19] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *The Adaptive Immune System*. 2002, ch. 24.

[20] P. Marrack, J. Kappler, and B. L. Kotzin, "Autoimmune disease: Why and where it occurs," *Nature Medicine*, vol. 7, no. 8, pp. 899–905, Aug. 2001.

[21] B. V. Kumar, T. Connors, and D. L. Farber, "Human T cell development, localization, and function throughout life," *Immunity*, vol. 48, no. 2, pp. 202–213, Feb. 2018.

[22] J. Charles A Janeway, P. Travers, M. Walport, and M. J. Shlomchik, "Immunological memory," *Immunobiology: The Immune System in Health and Disease. 5th edition*, 2001.

[23] S. Gallucci and P. Matzinger, "Danger signals: Sos to the immune system, current opinions in immunology 13," 2001.

[24] J. P. Anderson, "Computer security technology planning study," vol. 2, 1972.

[25] J. P. Anderson, "Computer Security Threat Monitoring and Surveillance," 1980.

[26] D. E. Denning, "An intrusion-detection model," in *1986 IEEE Symposium on Security and Privacy*, 1986, pp. 118–118.

[27] L. Heberlein, G. Dias, K. Levitt, B. Mukherjee, J. Wood, and D. Wolber, "A network security monitor," in *Proceedings. 1990 IEEE Computer Society Symposium on Research in Security and Privacy*, 1990, pp. 296–304.

[28] S. X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," en, *Applied Soft Computing*, vol. 10, no. 1, pp. 1–35, Jan. 2010.

[29] J. D. Farmer, N. H. Packard, and A. S. Perelson, "The immune system, adaptation, and machine learning," *Physica D: Nonlinear Phenomena*, vol. 22, no. 1-3, pp. 187–204, 1986.

[30] J. Tuo, S. Ren, W. Liu, X. Li, B. Li, and L. Lei, "Artificial immune system for fraud detection," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, vol. 2, 2004, 1407–1411 vol.2.

[31] L. de Castro and F. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239–251, 2002.

[32] A. Watkins, J. Timmis, and L. Boggess, "Artificial immune recognition system (airs): An immune-inspired supervised learning algorithm," *Genetic Programming and Evolvable Machines*, vol. 5, no. 3, pp. 291–317, 2004.

[33] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," in *Proceedings of 1994 IEEE computer society symposium on research in security and privacy*, Ieee, 1994, pp. 202–212.

[34] P. THAPA and D. L. FARBER, "THE ROLE OF THE THYMUS IN THE IMMUNE RESPONSE," *Thoracic surgery clinics*, vol. 29, no. 2, pp. 123–131, May 2019.

[35] J. Kim and P. J. Bentley, "An evaluation of negative selection in an artificial immune system for network intrusion detection," in *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, 2001, pp. 1330–1337.

[36] G. Costa Silva, R. M. Palhares, and W. M. Caminhas, "Immune inspired Fault Detection and Diagnosis: A fuzzy-based approach of the negative selection algorithm and participatory clustering," en, *Expert Systems with Applications*, vol. 39, no. 16, pp. 12 474–12 486, Nov. 2012.

[37] F. Gonzalez, D. Dasgupta, and R. Kozma, "Combining negative selection and classification techniques for anomaly detection," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, IEEE, vol. 1, 2002, pp. 705–710.

[38] S. F. M. Burnet *et al.*, *The clonal selection theory of acquired immunity*. Vanderbilt University Press Nashville, 1959, vol. 3.

[39] L. N. De Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE transactions on evolutionary computation*, vol. 6, no. 3, pp. 239–251, 2002.

[40] V. Cutello, G. Narzisi, G. Nicosia, and M. Pavone, "Clonal selection algorithms: A comparative case study using effective mutation potentials," in *International Conference on Artificial Immune Systems*, Springer, 2005, pp. 13–28.

[41] B. H. Ulutas and A. A. Islier, "A clonal selection algorithm for dynamic facility layout problems," *Journal of Manufacturing Systems*, vol. 28, no. 4, pp. 123–131, 2009.

[42] J. Timmis, M. Neal, and J. Hunt, "An artificial immune system for data analysis," *Biosystems*, vol. 55, no. 1, pp. 143–150, 2000.

[43] X. Shen, X. Gao, R. Bie, and X. Jin, "Artificial immune networks: Models and applications," vol. 1, Dec. 2006, pp. 394–397.

[44] U. Aickelin, P. Bentley, S. Cayzer, J. Kim, and J. McLeod, "Danger theory: The link between ais and ids?" In *Artificial Immune Systems*, J. Timmis, P. J. Bentley, and E. Hart, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 147–155.

[45] S. Rowan and D. Fisher, "Mechanisms of apoptotic cell death," *Leukemia*, vol. 11, no. 4, pp. 457–465, 1997.

[46] N. Vanlangenakker, T. V. Berghe, D. V. Krysko, N. Festjens, and P. Vandenabeele, “Molecular mechanisms and pathophysiology of necrotic cell death,” *Current molecular medicine*, vol. 8, no. 3, pp. 207–220, 2008.

[47] D. Xu and Y. Tian, “A comprehensive survey of clustering algorithms,” *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.

[48] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: A review,” *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[49] J. Page, Z. Liechty, M. Huynh, and J. Udall, “Bambam: Genome sequence analysis tools for biologists,” *BMC research notes*, vol. 7, p. 829, Nov. 2014.

[50] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek, “Density-based clustering,” *Wiley interdisciplinary reviews: data mining and knowledge discovery*, vol. 1, no. 3, pp. 231–240, 2011.

[51] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise.,” vol. 96, no. 34, pp. 226–231, 1996.

[52] S. Dang, “Performance evaluation of clustering algorithm using different datasets,” *IJARCSMS*, vol. 3, pp. 167–173, Jan. 2015.

[53] P. Pai, *Hierarchical clustering explained*, May 2021. [Online]. Available: `https://towardsdatascience.com/hierarchical-clustering-explained-e59b13846da8` (visited on Nov. 14, 2022).

[54] J. Saarikoski, J. Laurikkala, K. Järvelin, and M. Juhola, “Self-organising maps in document classification: A comparison with six machine learning methods,” in *Adaptive and Natural Computing Algorithms*, A. Dobnikar, U. Lotrič, and B. Šter, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 260–269.